

# CharsetDetector

CharsetDetector è un'applicazione in grado di riconoscere il charset di un determinato file.

In ingresso viene fornito il path di una directory, tutto il contenuto viene analizzato e stampata a video il tipo di codifica trovato.

Per la realizzazione è stata utilizzata la libreria juniversalchardet-1.0.3.jar che consiste in un port java della libreria di encoding detector di Mozilla.

Per verificarne il comportamento sono stati effettuati dei test con diversi tipi di file:

- Appunti AV-SOFT Ingegneria del Software.pdf  
Ci si aspetta un esito negativo ed infatti il programma restituisce "Encoding not recognized";
- C.csv  
E' un csv codificato opportunamente in windows-1252 (ANSI) ed effettivamente il programma riconosce la codifica e restituisce : "windows-1252";
- Le avventure di Pinocchio – Carlo Collodi.pdf  
Ci si aspetta un esito negativo ed infatti il programma restituisce "Encoding not recognized";
- prova.txt  
Testo con codifica UTF-8 ed il programma restituisce effettivamente "UTF-8";
- Text.txt  
File ottenuto convertendo in windows-1252 (ANSI) un .csv codificato in UTF-8. Il programma restituisce effettivamente "windows-1252".